Elaborato Modello Robusto

Marco Schiavi ; Edoardo Nada

## INTRODUZIONE

**ID** = Id riferito alla singola persona

**Diagnosis** = Diagnosi del tumore(M = maligno, B = benigno)

Per ciascun nucleo cellulare vengono calcolate dieci caratteristiche di valore reale

a) **radius** = media delle distanze dal centro ai punti del perimetro del nucleo cellulare.

b) **texture** = deviazione standard dai valori della scala dei grigi.

c) **perimeter** = perimetro del nucleo cellulare.

d) **area** = area del nucleo cellulare.

e) **smoothness** = variazione locale delle lunghezze del raggio.

f) **compactness** = Compattezza del nucleo cellulare.

g) **concavity** = Gravità delle porzioni concave del contorno del nucleo cellulare.

h) **concave points** = numero di porzioni concave del contorno del nucleo cellulare.

i) **symmetry**=simmetria del nucleo cellulare

j) **fractal dimension** = la dimensione frattale è un indicatore del grado di irregolarità del nucleo cellulare.

## OBBIETTIVO DEL MODELLO:

## Il modello mira a spiegare come alcune caratteristiche cellulari influenzano la dimensione frattale del nucleo della cellula. Questo studio è interessante poiché maggiore è la dimensione frattale di un nucleo maggiore sarà la sua tendenza ad essere associato ad un tumore Maligno.

## CONTROLLO MISSING VALUE

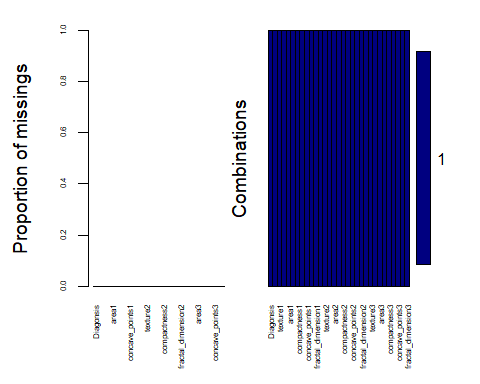
Dal controllo emerge che siamo in presenza di un dataset completo senza valori mancanti

library(datasets)  
library(VIM)

sapply(Tumori, function(x)(sum(is.na(x))))

## Diagonsis radius1 texture1 perimeter1   
## 0 0 0 0   
## area1 smoothness1 compactness1 concavity1   
## 0 0 0 0   
## concave\_points1 symmetry1 fractal\_dimension1 radius2   
## 0 0 0 0   
## texture2 perimeter2 area2 smoothness2   
## 0 0 0 0   
## compactness2 concavity2 concave\_points2 symmetry2   
## 0 0 0 0   
## fractal\_dimension2 radius3 texture3 perimeter3   
## 0 0 0 0   
## area3 smoothness3 compactness3 concavity3   
## 0 0 0 0   
## concave\_points3 symmetry3 fractal\_dimension3   
## 0 0 0

missingness<- aggr(Tumori, col=c('navyblue','yellow'),numbers=TRUE, sortVars=TRUE,labels=names(Tumori), cex.axis=.5,gap=2)



##   
## Variables sorted by number of missings:   
## Variable Count  
## Diagonsis 0  
## radius1 0  
## texture1 0  
## perimeter1 0  
## area1 0  
## smoothness1 0  
## compactness1 0  
## concavity1 0  
## concave\_points1 0  
## symmetry1 0  
## fractal\_dimension1 0  
## radius2 0  
## texture2 0  
## perimeter2 0  
## area2 0  
## smoothness2 0  
## compactness2 0  
## concavity2 0  
## concave\_points2 0  
## symmetry2 0  
## fractal\_dimension2 0  
## radius3 0  
## texture3 0  
## perimeter3 0  
## area3 0  
## smoothness3 0  
## compactness3 0  
## concavity3 0  
## concave\_points3 0  
## symmetry3 0  
## fractal\_dimension3 0

## COLLINEARITÀ

Partendo dal dataset completo creo un modello M1 che comprende tutte le variabili meno una e procedo alla creazione del miglior modello attraverso il comando step ‘SBC’ una volta creata M2 procedo ad un controllo manuale eliminando le variabili con un VIF>5 o TOL<0.3 e rifaccio i controlli per ogni singola variabile tolta fino a creare il modello M3 che non presenta collinearità infine faccio un plot grafico delle variabili per vedere anche graficamente la correlazione e il tipo di distribuzione

library(psych)  
library(corrgram)  
require(corrgram)  
library(mctest)  
  
M1<-lm(fractal\_dimension1~ .-Diagonsis , data=Tumori)  
  
selectedMod <- step(M1, direction="both", k = log(nrow(Tumori)))

M2<-lm(fractal\_dimension1 ~ perimeter1 + area1 + smoothness1 + compactness1 +   
 perimeter2 + area2 + concavity2 + compactness3 + concavity3 +   
 fractal\_dimension3, data=Tumori)  
imcdiag(M2)

##   
## Call:  
## imcdiag(mod = M2)  
##   
##   
## All Individual Multicollinearity Diagnostics Result  
##   
## VIF TOL Wi Fi Leamer CVIF Klein  
## perimeter1 82.4262 0.0121 5057.4703 5699.8324 0.1101 -5.9869 1  
## area1 88.3659 0.0113 5426.3948 6115.6148 0.1064 -6.4184 1  
## smoothness1 2.5090 0.3986 93.7271 105.6316 0.6313 -0.1822 0  
## compactness1 12.3708 0.0808 706.2530 795.9559 0.2843 -0.8985 0  
## perimeter2 12.2410 0.0817 698.1906 786.8695 0.2858 -0.8891 0  
## area2 16.4246 0.0609 958.0381 1079.7209 0.2467 -1.1930 1  
## concavity2 3.2172 0.3108 137.7099 155.2007 0.5575 -0.2337 0  
## compactness3 15.0951 0.0662 875.4633 986.6581 0.2574 -1.0964 1  
## concavity3 11.2171 0.0891 634.5938 715.1951 0.2986 -0.8147 0  
## fractal\_dimension3 5.4417 0.1838 275.8787 310.9187 0.4287 -0.3953 0  
## IND1 IND2  
## perimeter1 0.0002 1.1349  
## area1 0.0002 1.1358  
## smoothness1 0.0064 0.6909  
## compactness1 0.0013 1.0560  
## perimeter2 0.0013 1.0550  
## area2 0.0010 1.0789  
## concavity2 0.0050 0.7917  
## compactness3 0.0011 1.0727  
## concavity3 0.0014 1.0464  
## fractal\_dimension3 0.0030 0.9377  
##   
## 1 --> COLLINEARITY is detected by the test   
## 0 --> COLLINEARITY is not detected by the test  
##   
## \* all coefficients have significant t-ratios  
##   
## R-square of y on all x: 0.9318   
##   
## \* use method argument to check which regressors may be the reason of collinearity  
## ===================================

summary(M2)

##   
## Call:  
## lm(formula = fractal\_dimension1 ~ perimeter1 + area1 + smoothness1 +   
## compactness1 + perimeter2 + area2 + concavity2 + compactness3 +   
## concavity3 + fractal\_dimension3, data = Tumori)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.0066263 -0.0011583 -0.0000357 0.0010991 0.0075094   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.055346248 0.001779306 31.106 < 0.0000000000000002 \*\*\*  
## perimeter1 -0.000389788 0.000029156 -13.369 < 0.0000000000000002 \*\*\*  
## area1 0.000018059 0.000002084 8.664 < 0.0000000000000002 \*\*\*  
## smoothness1 0.040155762 0.008788733 4.569 0.0000060369 \*\*\*  
## compactness1 0.122757542 0.005196929 23.621 < 0.0000000000000002 \*\*\*  
## perimeter2 0.000443976 0.000135035 3.288 0.00107 \*\*   
## area2 -0.000030068 0.000006952 -4.325 0.0000180633 \*\*\*  
## concavity2 0.025092367 0.004636777 5.412 0.0000000929 \*\*\*  
## compactness3 -0.024919456 0.001926973 -12.932 < 0.0000000000000002 \*\*\*  
## concavity3 -0.005089147 0.001252742 -4.062 0.0000555311 \*\*\*  
## fractal\_dimension3 0.257980572 0.010078728 25.597 < 0.0000000000000002 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.00186 on 558 degrees of freedom  
## Multiple R-squared: 0.9318, Adjusted R-squared: 0.9306   
## F-statistic: 762.8 on 10 and 558 DF, p-value: < 0.00000000000000022

drop1(M2, test="F")

## Single term deletions  
##   
## Model:  
## fractal\_dimension1 ~ perimeter1 + area1 + smoothness1 + compactness1 +   
## perimeter2 + area2 + concavity2 + compactness3 + concavity3 +   
## fractal\_dimension3  
## Df Sum of Sq RSS AIC F value  
## <none> 0.0019300 -7144.1   
## perimeter1 1 0.00061818 0.0025482 -6988.0 178.729  
## area1 1 0.00025962 0.0021896 -7074.2 75.061  
## smoothness1 1 0.00007220 0.0020022 -7125.2 20.876  
## compactness1 1 0.00192985 0.0038598 -6751.7 557.960  
## perimeter2 1 0.00003739 0.0019674 -7135.1 10.810  
## area2 1 0.00006470 0.0019947 -7127.3 18.706  
## concavity2 1 0.00010129 0.0020313 -7116.9 29.285  
## compactness3 1 0.00057843 0.0025084 -6996.9 167.235  
## concavity3 1 0.00005708 0.0019871 -7129.5 16.503  
## fractal\_dimension3 1 0.00226613 0.0041961 -6704.1 655.183  
## Pr(>F)   
## <none>   
## perimeter1 < 0.00000000000000022 \*\*\*  
## area1 < 0.00000000000000022 \*\*\*  
## smoothness1 0.00000603686 \*\*\*  
## compactness1 < 0.00000000000000022 \*\*\*  
## perimeter2 0.001073 \*\*   
## area2 0.00001806334 \*\*\*  
## concavity2 0.00000009286 \*\*\*  
## compactness3 < 0.00000000000000022 \*\*\*  
## concavity3 0.00005553106 \*\*\*  
## fractal\_dimension3 < 0.00000000000000022 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

M3<-lm(fractal\_dimension1 ~ smoothness1 + area2 + concavity2 + concavity3 +   
 fractal\_dimension3, data=Tumori)  
  
  
imcdiag(M3)

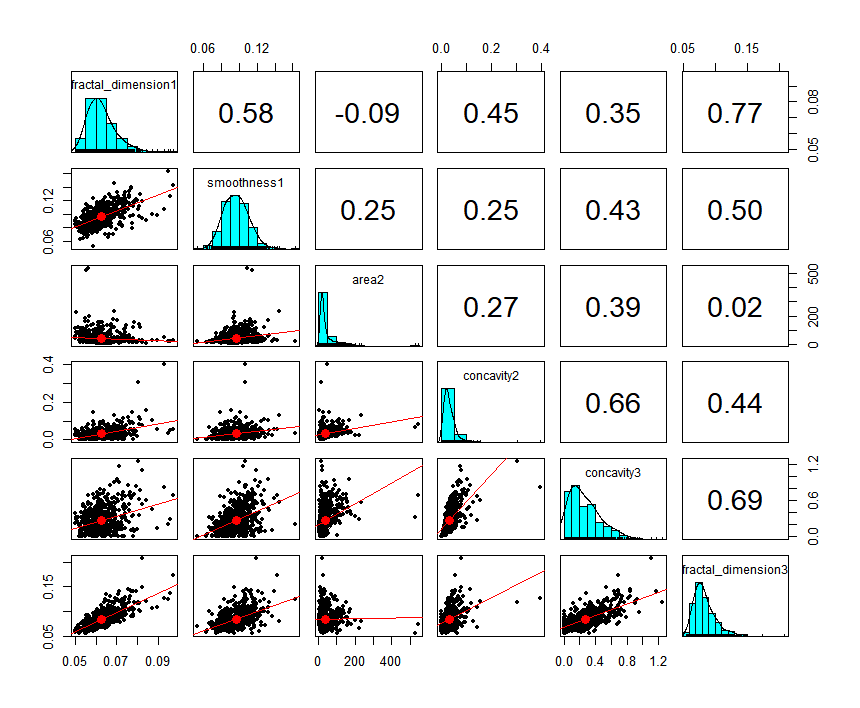
##   
## Call:  
## imcdiag(mod = M3)  
##   
##   
## All Individual Multicollinearity Diagnostics Result  
##   
## VIF TOL Wi Fi Leamer CVIF Klein IND1  
## smoothness1 1.4455 0.6918 62.8178 83.9056 0.8317 -0.9373 0 0.0049  
## area2 1.4382 0.6953 61.7798 82.5191 0.8339 -0.9325 0 0.0049  
## concavity2 1.7902 0.5586 111.4161 148.8182 0.7474 -1.1608 0 0.0040  
## concavity3 3.3833 0.2956 336.0520 448.8638 0.5437 -2.1938 0 0.0021  
## fractal\_dimension3 2.5322 0.3949 216.0397 288.5637 0.6284 -1.6419 0 0.0028  
## IND2  
## smoothness1 0.6519  
## area2 0.6444  
## concavity2 0.9337  
## concavity3 1.4901  
## fractal\_dimension3 1.2799  
##   
## 1 --> COLLINEARITY is detected by the test   
## 0 --> COLLINEARITY is not detected by the test  
##   
## \* all coefficients have significant t-ratios  
##   
## R-square of y on all x: 0.8326   
##   
## \* use method argument to check which regressors may be the reason of collinearity  
## ===================================

summary(M3)

##   
## Call:  
## lm(formula = fractal\_dimension1 ~ smoothness1 + area2 + concavity2 +   
## concavity3 + fractal\_dimension3, data = Tumori)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.0117470 -0.0016636 -0.0002098 0.0014497 0.0136872   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.020944093 0.000969046 21.61 < 0.0000000000000002 \*\*\*  
## smoothness1 0.175927894 0.010408548 16.90 < 0.0000000000000002 \*\*\*  
## area2 -0.000008538 0.000003210 -2.66 0.00804 \*\*   
## concavity2 0.102278678 0.005396772 18.95 < 0.0000000000000002 \*\*\*  
## concavity3 -0.022412250 0.001073495 -20.88 < 0.0000000000000002 \*\*\*  
## fractal\_dimension3 0.334546076 0.010727339 31.19 < 0.0000000000000002 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.002902 on 563 degrees of freedom  
## Multiple R-squared: 0.8326, Adjusted R-squared: 0.8311   
## F-statistic: 559.9 on 5 and 563 DF, p-value: < 0.00000000000000022

drop1(M3, test="F")

## Single term deletions  
##   
## Model:  
## fractal\_dimension1 ~ smoothness1 + area2 + concavity2 + concavity3 +   
## fractal\_dimension3  
## Df Sum of Sq RSS AIC F value Pr(>F)  
## <none> 0.0047407 -6642.7   
## smoothness1 1 0.0024056 0.0071463 -6411.2 285.686 < 0.00000000000000022  
## area2 1 0.0000596 0.0048002 -6637.6 7.075 0.008039  
## concavity2 1 0.0030244 0.0077650 -6363.9 359.172 < 0.00000000000000022  
## concavity3 1 0.0036703 0.0084110 -6318.5 435.884 < 0.00000000000000022  
## fractal\_dimension3 1 0.0081895 0.0129302 -6073.8 972.586 < 0.00000000000000022  
##   
## <none>   
## smoothness1 \*\*\*  
## area2 \*\*   
## concavity2 \*\*\*  
## concavity3 \*\*\*  
## fractal\_dimension3 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

  
Corr

## fractal\_dimension1 smoothness1 area2 concavity2 concavity3  
## fractal\_dimension1 1.000 0.585 -0.090 0.447 0.346  
## smoothness1 0.585 1.000 0.247 0.248 0.435  
## area2 -0.090 0.247 1.000 0.271 0.385  
## concavity2 0.447 0.248 0.271 1.000 0.663  
## concavity3 0.346 0.435 0.385 0.663 1.000  
## fractal\_dimension3 0.767 0.499 0.018 0.439 0.687  
## fractal\_dimension3  
## fractal\_dimension1 0.767  
## smoothness1 0.499  
## area2 0.018  
## concavity2 0.439  
## concavity3 0.687  
## fractal\_dimension3 1.000

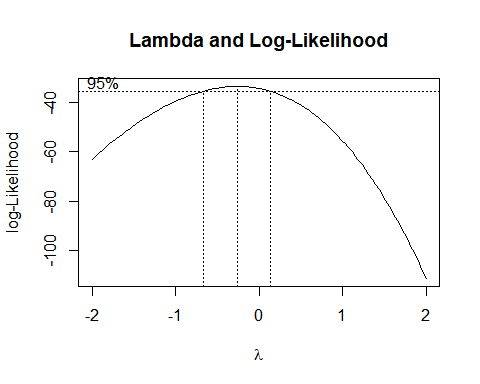
## LINEARITÀ

Dal test BOX-COX notiamo un valore lambda tale da farci trattare la Y con un logaritmo.

Attraverso la funzione gam e il plot possiamo vedere una significativtà in tre variabili ma decidiamo di modificarne solo due di esse essendo le x disposte in maniera tale da rendere piu logica la modifica delle relative variabili per la linearizzazione ( area2, concavity2) Dal resettest si evidenzia un netto miglioramento

library(MASS)  
library(dplyr)

library(stats)  
library(base)  
boxcoxreg1<-boxcox(M3)  
title("Lambda and Log-Likelihood")



lambda=boxcoxreg1$x[which.max(boxcoxreg1$y)]  
lambda

## [1] -0.2626263

M4<-lm((I(fractal\_dimension1)) ~ smoothness1 + area2 + concavity2 + concavity3 +   
 fractal\_dimension3, data=Tumori)  
  
  
summary(M3)

##   
## Call:  
## lm(formula = fractal\_dimension1 ~ smoothness1 + area2 + concavity2 +   
## concavity3 + fractal\_dimension3, data = Tumori)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.0117470 -0.0016636 -0.0002098 0.0014497 0.0136872   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.020944093 0.000969046 21.61 < 0.0000000000000002 \*\*\*  
## smoothness1 0.175927894 0.010408548 16.90 < 0.0000000000000002 \*\*\*  
## area2 -0.000008538 0.000003210 -2.66 0.00804 \*\*   
## concavity2 0.102278678 0.005396772 18.95 < 0.0000000000000002 \*\*\*  
## concavity3 -0.022412250 0.001073495 -20.88 < 0.0000000000000002 \*\*\*  
## fractal\_dimension3 0.334546076 0.010727339 31.19 < 0.0000000000000002 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.002902 on 563 degrees of freedom  
## Multiple R-squared: 0.8326, Adjusted R-squared: 0.8311   
## F-statistic: 559.9 on 5 and 563 DF, p-value: < 0.00000000000000022

summary(M4)

##   
## Call:  
## lm(formula = (I(fractal\_dimension1)) ~ smoothness1 + area2 +   
## concavity2 + concavity3 + fractal\_dimension3, data = Tumori)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.0117470 -0.0016636 -0.0002098 0.0014497 0.0136872   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.020944093 0.000969046 21.61 < 0.0000000000000002 \*\*\*  
## smoothness1 0.175927894 0.010408548 16.90 < 0.0000000000000002 \*\*\*  
## area2 -0.000008538 0.000003210 -2.66 0.00804 \*\*   
## concavity2 0.102278678 0.005396772 18.95 < 0.0000000000000002 \*\*\*  
## concavity3 -0.022412250 0.001073495 -20.88 < 0.0000000000000002 \*\*\*  
## fractal\_dimension3 0.334546076 0.010727339 31.19 < 0.0000000000000002 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.002902 on 563 degrees of freedom  
## Multiple R-squared: 0.8326, Adjusted R-squared: 0.8311   
## F-statistic: 559.9 on 5 and 563 DF, p-value: < 0.00000000000000022

drop1(M3, test="F")

## Single term deletions  
##   
## Model:  
## fractal\_dimension1 ~ smoothness1 + area2 + concavity2 + concavity3 +   
## fractal\_dimension3  
## Df Sum of Sq RSS AIC F value Pr(>F)  
## <none> 0.0047407 -6642.7   
## smoothness1 1 0.0024056 0.0071463 -6411.2 285.686 < 0.00000000000000022  
## area2 1 0.0000596 0.0048002 -6637.6 7.075 0.008039  
## concavity2 1 0.0030244 0.0077650 -6363.9 359.172 < 0.00000000000000022  
## concavity3 1 0.0036703 0.0084110 -6318.5 435.884 < 0.00000000000000022  
## fractal\_dimension3 1 0.0081895 0.0129302 -6073.8 972.586 < 0.00000000000000022  
##   
## <none>   
## smoothness1 \*\*\*  
## area2 \*\*   
## concavity2 \*\*\*  
## concavity3 \*\*\*  
## fractal\_dimension3 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

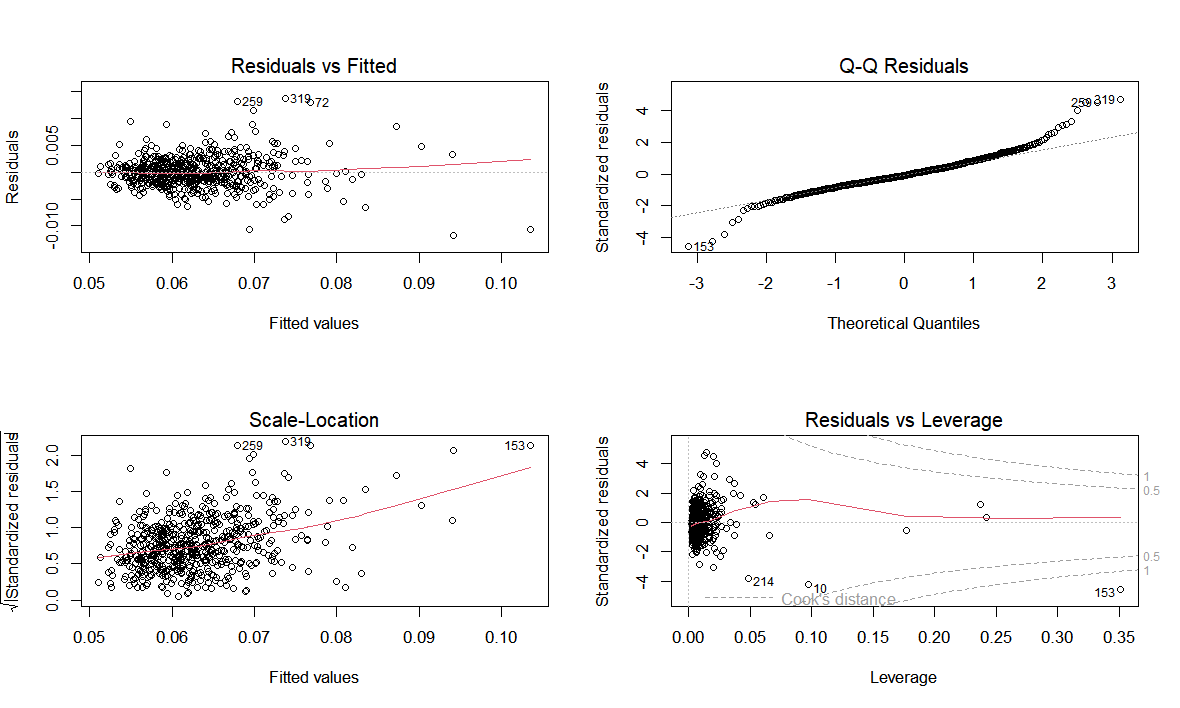
drop1(M4, test="F")

## Single term deletions  
##   
## Model:  
## (I(fractal\_dimension1)) ~ smoothness1 + area2 + concavity2 +   
## concavity3 + fractal\_dimension3  
## Df Sum of Sq RSS AIC F value Pr(>F)  
## <none> 0.0047407 -6642.7   
## smoothness1 1 0.0024056 0.0071463 -6411.2 285.686 < 0.00000000000000022  
## area2 1 0.0000596 0.0048002 -6637.6 7.075 0.008039  
## concavity2 1 0.0030244 0.0077650 -6363.9 359.172 < 0.00000000000000022  
## concavity3 1 0.0036703 0.0084110 -6318.5 435.884 < 0.00000000000000022  
## fractal\_dimension3 1 0.0081895 0.0129302 -6073.8 972.586 < 0.00000000000000022  
##   
## <none>   
## smoothness1 \*\*\*  
## area2 \*\*   
## concavity2 \*\*\*  
## concavity3 \*\*\*  
## fractal\_dimension3 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

par(mfrow=c(2,2))   
plot(M3)

par(mfrow=c(1,1))  
par(mfrow=c(2,2))   
plot(M4)  
par(mfrow=c(1,1))

# Tra M3 e M4 a livello grafico non cambia nulla inserisco quindi un solo plot grafico

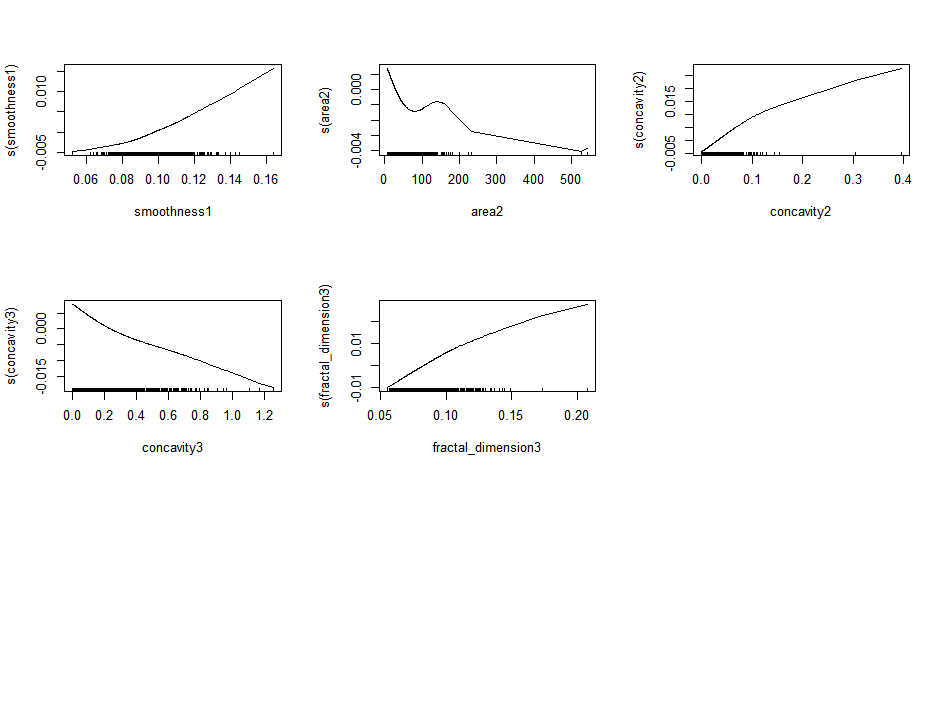


library(splines)  
library(foreach)  
library(base)  
library(gam)

m4gam<- gam((I(fractal\_dimension1)) ~ s(smoothness1) + s(area2) + s(concavity2) + s(concavity3) +s(fractal\_dimension3), data=Tumori)  
summary(m4gam)

##   
## Call: gam(formula = (I(fractal\_dimension1)) ~ s(smoothness1) + s(area2) +   
## s(concavity2) + s(concavity3) + s(fractal\_dimension3), data = Tumori)  
## Deviance Residuals:  
## Min 1Q Median 3Q Max   
## -0.0098522 -0.0013890 -0.0001462 0.0012478 0.0122680   
##   
## (Dispersion Parameter for gaussian family taken to be 0)  
##   
## Null Deviance: 0.0283 on 568 degrees of freedom  
## Residual Deviance: 0.0035 on 548.0001 degrees of freedom  
## AIC: -5165.112   
##   
## Number of Local Scoring Iterations: NA   
##   
## Anova for Parametric Effects  
## Df Sum Sq Mean Sq F value Pr(>F)  
## s(smoothness1) 1 0.0090746 0.0090746 1412.123 < 0.00000000000000022  
## s(area2) 1 0.0019556 0.0019556 304.311 < 0.00000000000000022  
## s(concavity2) 1 0.0043420 0.0043420 675.677 < 0.00000000000000022  
## s(concavity3) 1 0.0001550 0.0001550 24.121 0.000001196  
## s(fractal\_dimension3) 1 0.0070068 0.0070068 1090.344 < 0.00000000000000022  
## Residuals 548 0.0035216 0.0000064   
##   
## s(smoothness1) \*\*\*  
## s(area2) \*\*\*  
## s(concavity2) \*\*\*  
## s(concavity3) \*\*\*  
## s(fractal\_dimension3) \*\*\*  
## Residuals   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Anova for Nonparametric Effects  
## Npar Df Npar F Pr(F)   
## (Intercept)   
## s(smoothness1) 3 7.0692 0.0001144 \*\*\*  
## s(area2) 3 11.3584 0.000000308588314235 \*\*\*  
## s(concavity2) 3 16.4710 0.000000000292609936 \*\*\*  
## s(concavity3) 3 24.7144 0.000000000000005218 \*\*\*  
## s(fractal\_dimension3) 3 17.8318 0.000000000046825321 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

par(mfrow=c(3,3))   
plot(m4gam)  
par(mfrow=c(1,1))



m4gam = lm((I(fractal\_dimension1)) ~ smoothness1 +area2+ (area2)^2 +concavity2+ I((concavity2)^2) + concavity3 +fractal\_dimension3, data=Tumori)  
summary(m4gam)

##   
## Call:  
## lm(formula = (I(fractal\_dimension1)) ~ smoothness1 + area2 +   
## (area2)^2 + concavity2 + I((concavity2)^2) + concavity3 +   
## fractal\_dimension3, data = Tumori)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.012347 -0.001636 -0.000177 0.001442 0.012634   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.020627391 0.000956243 21.571 < 0.0000000000000002 \*\*\*  
## smoothness1 0.175939509 0.010242030 17.178 < 0.0000000000000002 \*\*\*  
## area2 -0.000010601 0.000003193 -3.320 0.000958 \*\*\*  
## concavity2 0.139475352 0.009965709 13.996 < 0.0000000000000002 \*\*\*  
## I((concavity2)^2) -0.133623019 0.030294009 -4.411 0.0000123 \*\*\*  
## concavity3 -0.023923300 0.001110482 -21.543 < 0.0000000000000002 \*\*\*  
## fractal\_dimension3 0.333131312 0.010560591 31.545 < 0.0000000000000002 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.002855 on 562 degrees of freedom  
## Multiple R-squared: 0.8382, Adjusted R-squared: 0.8364   
## F-statistic: 485.1 on 6 and 562 DF, p-value: < 0.00000000000000022

drop1(m4gam, test="F")

## Single term deletions  
##   
## Model:  
## (I(fractal\_dimension1)) ~ smoothness1 + area2 + (area2)^2 + concavity2 +   
## I((concavity2)^2) + concavity3 + fractal\_dimension3  
## Df Sum of Sq RSS AIC F value Pr(>F)  
## <none> 0.0045820 -6660.1   
## smoothness1 1 0.0024059 0.0069879 -6421.9 295.090 < 0.00000000000000022  
## area2 1 0.0000899 0.0046719 -6651.0 11.024 0.0009576  
## concavity2 1 0.0015970 0.0061790 -6491.9 195.875 < 0.00000000000000022  
## I((concavity2)^2) 1 0.0001586 0.0047407 -6642.7 19.456 0.00001234  
## concavity3 1 0.0037839 0.0083660 -6319.5 464.108 < 0.00000000000000022  
## fractal\_dimension3 1 0.0081129 0.0126950 -6082.2 995.072 < 0.00000000000000022  
##   
## <none>   
## smoothness1 \*\*\*  
## area2 \*\*\*  
## concavity2 \*\*\*  
## I((concavity2)^2) \*\*\*  
## concavity3 \*\*\*  
## fractal\_dimension3 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

library(base)  
library(zoo)

library(lmtest)  
  
#CONFRONTIAMO CON IL MODELLO FINALE CON I MODELLI INTERMEDI:  
resettest(M3, power = 2, type = "fitted", data = Tumori)

##   
## RESET test  
##   
## data: M3  
## RESET = 9.1773, df1 = 1, df2 = 562, p-value = 0.002563

resettest(M4, power = 2, type = "fitted", data = Tumori)

##   
## RESET test  
##   
## data: M4  
## RESET = 9.1773, df1 = 1, df2 = 562, p-value = 0.002563

resettest(m4gam, power = 2, type = "fitted", data = Tumori)

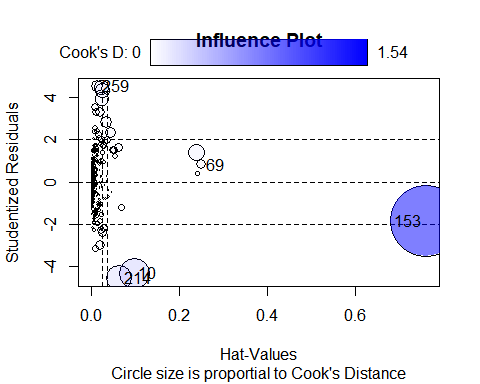
##   
## RESET test  
##   
## data: m4gam  
## RESET = 0.16308, df1 = 1, df2 = 561, p-value = 0.6865

## LEVERAGE E OUTLIER

Attraverso l’analisi dei leverage e degli outlier procediamo alla rimozione dei punti influenti.

library(carData)  
library(psych)  
library(dplyr)  
library(car)

influencePlot(m4gam, main="Influence Plot", sub="Circle size is proportial to Cook's Distance" )

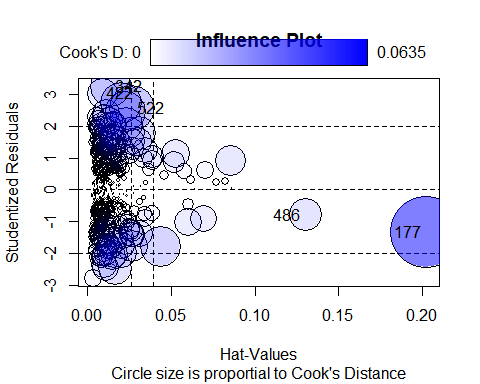


## StudRes Hat CookD  
## 10 -4.3043331 0.09750266 0.27729795  
## 69 0.8230228 0.25139877 0.03251529  
## 153 -1.8414546 0.76129236 1.53838633  
## 214 -4.5498677 0.06508960 0.19891951  
## 259 4.5330252 0.01410566 0.04058746

res=data.frame(m4gam$residuals)  
data\_used=Tumori[rownames(res),]  
  
  
cooksd <- cooks.distance(m4gam)  
cd=data.frame(cooksd)  
  
cutoff= 0.0023

NONinfluenti=data.frame(Tumori[cooksd < cutoff, ])   
  
  
  
m4Cook = lm((I(fractal\_dimension1)) ~ smoothness1 +area2+ (area2)^2 +concavity2+ I((concavity2)^2) + concavity3 +fractal\_dimension3, data=NONinfluenti)

influencePlot(m4Cook, main="Influence Plot", sub="Circle size is proportial to Cook's Distance" )



## StudRes Hat CookD

##177 -1.5736225 0.270998984 0.131093330

##183 -3.2985305 0.004215674 0.006444803

##273 1.8859141 0.038916971 0.020462931

##382 2.7149044 0.006087640 0.006363041

##486 -0.7427836 0.177832411 0.017064421

summary(m4gam)

##   
## Call:  
## lm(formula = (I(fractal\_dimension1)) ~ smoothness1 + area2 +   
## (area2)^2 + concavity2 + I((concavity2)^2) + concavity3 +   
## fractal\_dimension3, data = Tumori)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.012347 -0.001636 -0.000177 0.001442 0.012634   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.020627391 0.000956243 21.571 < 0.0000000000000002 \*\*\*  
## smoothness1 0.175939509 0.010242030 17.178 < 0.0000000000000002 \*\*\*  
## area2 -0.000010601 0.000003193 -3.320 0.000958 \*\*\*  
## concavity2 0.139475352 0.009965709 13.996 < 0.0000000000000002 \*\*\*  
## I((concavity2)^2) -0.133623019 0.030294009 -4.411 0.0000123 \*\*\*  
## concavity3 -0.023923300 0.001110482 -21.543 < 0.0000000000000002 \*\*\*  
## fractal\_dimension3 0.333131312 0.010560591 31.545 < 0.0000000000000002 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.002855 on 562 degrees of freedom  
## Multiple R-squared: 0.8382, Adjusted R-squared: 0.8364   
## F-statistic: 485.1 on 6 and 562 DF, p-value: < 0.00000000000000022

summary(m4Cook)

##   
## Call:

##lm(formula = (I(fractal\_dimension1)) ~ smoothness1 + area2 +

## (area2)^2 + concavity2 + I((concavity2)^2) + concavity3 +

## fractal\_dimension3, data = NONinfluenti)

##Residuals:

## Min 1Q Median 3Q Max

##-0.0060180 -0.0013132 -0.0001222 0.0012247 0.0049667

##Coefficients:

## Estimate Std. Error t value Pr(>|t|)

##(Intercept) 0.020774738 0.000767460 27.069 < 0.0000000000000002 \*\*\*

##smoothness1 0.180054165 0.008081170 22.281 < 0.0000000000000002 \*\*\*

##area2 -0.000016506 0.000003514 -4.698 0.000003459640 \*\*\*

##concavity2 0.105791820 0.016021059 6.603 0.000000000109 \*\*\*

##I((concavity2)^2) 0.181603430 0.157112225 1.156 0.248

##concavity3 -0.023242610 0.000974992 -23.839 < 0.0000000000000002 \*\*\*

##fractal\_dimension3 0.332966696 0.008952694 37.192 < 0.0000000000000002 \*\*\*

##---

##Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

##Residual standard error: 0.001847 on 470 degrees of freedom

##Multiple R-squared: 0.8895, Adjusted R-squared: 0.8881

##F-statistic: 630.5 on 6 and 470 DF, p-value: < 0.00000000000000022

drop1(m4gam, test="F")

## Single term deletions  
##   
## Model:  
## (I(fractal\_dimension1)) ~ smoothness1 + area2 + (area2)^2 + concavity2 +   
## I((concavity2)^2) + concavity3 + fractal\_dimension3  
## Df Sum of Sq RSS AIC F value Pr(>F)  
## <none> 0.0045820 -6660.1   
## smoothness1 1 0.0024059 0.0069879 -6421.9 295.090 < 0.00000000000000022  
## area2 1 0.0000899 0.0046719 -6651.0 11.024 0.0009576  
## concavity2 1 0.0015970 0.0061790 -6491.9 195.875 < 0.00000000000000022  
## I((concavity2)^2) 1 0.0001586 0.0047407 -6642.7 19.456 0.00001234  
## concavity3 1 0.0037839 0.0083660 -6319.5 464.108 < 0.00000000000000022  
## fractal\_dimension3 1 0.0081129 0.0126950 -6082.2 995.072 < 0.00000000000000022  
##   
## <none>   
## smoothness1 \*\*\*  
## area2 \*\*\*  
## concavity2 \*\*\*  
## I((concavity2)^2) \*\*\*  
## concavity3 \*\*\*  
## fractal\_dimension3 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

drop1(m4Cook, test="F")

## Single term deletions

##

##Model:

##(I(fractal\_dimension1)) ~ smoothness1 + area2 + (area2)^2 + concavity2 +

## I((concavity2)^2) + concavity3 + fractal\_dimension3

## Df Sum of Sq RSS AIC F value Pr(>F)

##<none> 0.0016041 -5997.5

##smoothness1 1 0.0016943 0.0032984 -5655.6 496.4298 < 0.00000000000000022

##area2 1 0.0000753 0.0016794 -5977.6 22.0684 0.0000034596400

##concavity2 1 0.0001488 0.0017529 -5957.2 43.6035 0.0000000001089

##I((concavity2)^2) 1 0.0000046 0.0016086 -5998.1 1.3361 0.2483

##concavity3 1 0.0019395 0.0035436 -5621.4 568.2875 < 0.00000000000000022

##fractal\_dimension3 1 0.0047209 0.0063250 -5345.1 1383.2291 < 0.00000000000000022

##<none>

##smoothness1 \*\*\*

##area2 \*\*\*

##concavity2 \*\*\*

##I((concavity2)^2)

##concavity3 \*\*\*

##fractal\_dimension3 \*\*\*

##---

##Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

resettest(m4gam, power = 2, type = "fitted", data = Tumori)

##   
## RESET test  
##   
## data: m4gam  
## RESET = 0.16308, df1 = 1, df2 = 561, p-value = 0.6865

resettest(m4Cook, power = 2, type = "fitted", data = NONinfluenti)

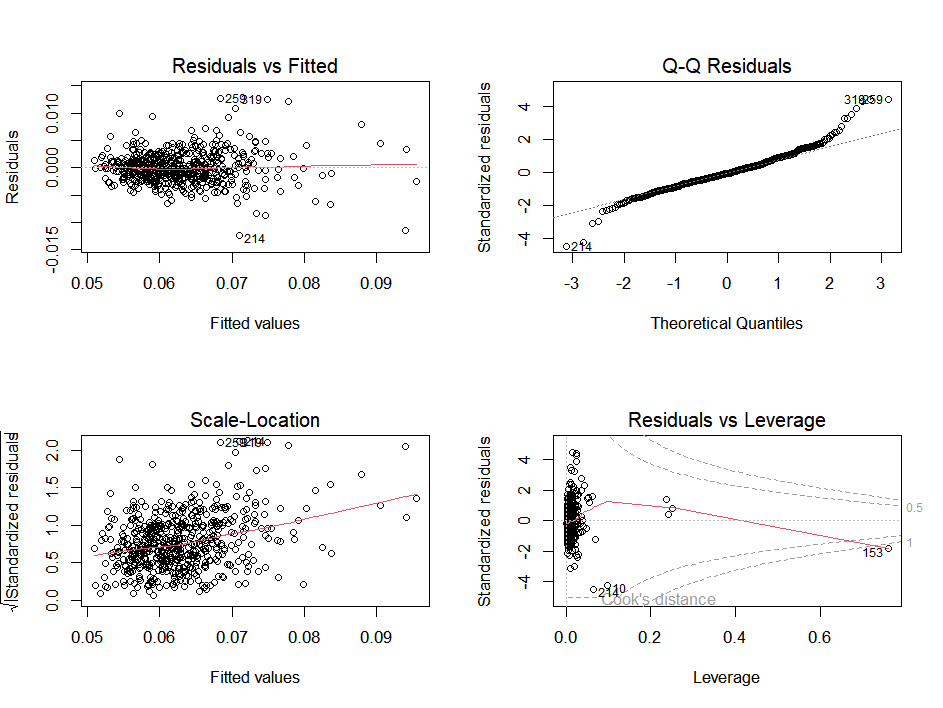
##   
## RESET test

##

##data: m4Cook

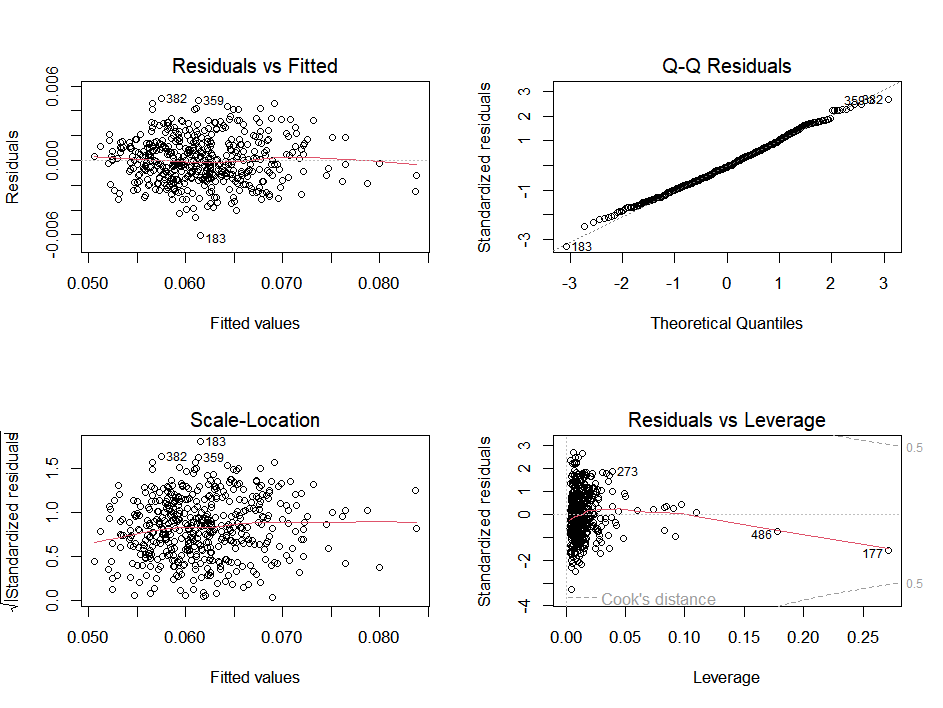
##RESET = 0.0067525, df1 = 1, df2 = 469, p-value = 0.9345

par(mfrow=c(2,2))  
plot(m4gam)



par(mfrow=c(1,1))  
  
par(mfrow=c(2,2))  
plot(m4Cook)

par(mfrow=c(1,1))



## RIMOZIONE VARIABILE

Finita la rimozione degli outlier e leverage vedo che la variabile I((concavity2)^2) non è più significativa procediamo quindi alla rimozione della stessa ed al successivo controllo del modello.

m5 = lm((I(fractal\_dimension1)) ~ smoothness1 +area2+ (area2)^2 +concavity2+ concavity3 +fractal\_dimension3, data=NONinfluenti)  
summary(m4Cook)

##   
## Call:

##lm(formula = (I(fractal\_dimension1)) ~ smoothness1 + area2 +

## (area2)^2 + concavity2 + I((concavity2)^2) + concavity3 +

## fractal\_dimension3, data = NONinfluenti)

##

##Residuals:

## Min 1Q Median 3Q Max

##-0.0060180 -0.0013132 -0.0001222 0.0012247 0.0049667

##

##Coefficients:

## Estimate Std. Error t value Pr(>|t|)

##(Intercept) 0.020774738 0.000767460 27.069 < 0.0000000000000002 \*\*\*

##smoothness1 0.180054165 0.008081170 22.281 < 0.0000000000000002 \*\*\*

##area2 -0.000016506 0.000003514 -4.698 0.000003459640 \*\*\*

##concavity2 0.105791820 0.016021059 6.603 0.000000000109 \*\*\*

##I((concavity2)^2) 0.181603430 0.157112225 1.156 0.248

##concavity3 -0.023242610 0.000974992 -23.839 < 0.0000000000000002 \*\*\*

##fractal\_dimension3 0.332966696 0.008952694 37.192 < 0.0000000000000002 \*\*\*

##---

##Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

##

##Residual standard error: 0.001847 on 470 degrees of freedom

##Multiple R-squared: 0.8895, Adjusted R-squared: 0.8881

##F-statistic: 630.5 on 6 and 470 DF, p-value: < 0.00000000000000022

summary(m5)

##   
## Call:

##lm(formula = (I(fractal\_dimension1)) ~ smoothness1 + area2 +

## (area2)^2 + concavity2 + concavity3 + fractal\_dimension3,

## data = NONinfluenti)

##

##Residuals:

## Min 1Q Median 3Q Max

##-0.0060040 -0.0013010 -0.0001329 0.0012093 0.0049505

##

##Coefficients:

## Estimate Std. Error t value Pr(>|t|)

##(Intercept) 0.020632492 0.000757800 27.227 < 0.0000000000000002 \*\*\*

##smoothness1 0.178772469 0.008007590 22.325 < 0.0000000000000002 \*\*\*

##area2 -0.000016373 0.000003513 -4.661 0.00000411 \*\*\*

##concavity2 0.122625870 0.006678475 18.361 < 0.0000000000000002 \*\*\*

##concavity3 -0.023629712 0.000915997 -25.797 < 0.0000000000000002 \*\*\*

##fractal\_dimension3 0.334109012 0.008901155 37.535 < 0.0000000000000002 \*\*\*

##---

##Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

##

##Residual standard error: 0.001848 on 471 degrees of freedom

##Multiple R-squared: 0.8892, Adjusted R-squared: 0.888

##F-statistic: 755.8 on 5 and 471 DF, p-value: < 0.00000000000000022

drop1(m4Cook, test="F")

## Single term deletions

##Model:

##(I(fractal\_dimension1)) ~ smoothness1 + area2 + (area2)^2 + concavity2 +

## I((concavity2)^2) + concavity3 + fractal\_dimension3

## Df Sum of Sq RSS AIC F value Pr(>F)

##<none> 0.0016041 -5997.5

##smoothness1 1 0.0016943 0.0032984 -5655.6 496.4298 < 0.00000000000000022

##area2 1 0.0000753 0.0016794 -5977.6 22.0684 0.0000034596400

##concavity2 1 0.0001488 0.0017529 -5957.2 43.6035 0.0000000001089

##I((concavity2)^2) 1 0.0000046 0.0016086 -5998.1 1.3361 0.2483

##concavity3 1 0.0019395 0.0035436 -5621.4 568.2875 < 0.00000000000000022

##fractal\_dimension3 1 0.0047209 0.0063250 -5345.1 1383.2291 < 0.00000000000000022

##<none>

##smoothness1 \*\*\*

##area2 \*\*\*

##concavity2 \*\*\*

##I((concavity2)^2)

##concavity3 \*\*\*

##fractal\_dimension3 \*\*\*

##---

##Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

drop1(m5, test="F")

## Single term deletions

##Model:

##(I(fractal\_dimension1)) ~ smoothness1 + area2 + (area2)^2 + concavity2 +

## concavity3 + fractal\_dimension3

## Df Sum of Sq RSS AIC F value Pr(>F)

##<none> 0.0016086 -5998.1

##smoothness1 1 0.0017023 0.0033109 -5655.8 498.423 < 0.00000000000000022

##area2 1 0.0000742 0.0016828 -5978.6 21.722 0.000004108

##concavity2 1 0.0011515 0.0027601 -5742.6 337.139 < 0.00000000000000022

##concavity3 1 0.0022728 0.0038815 -5580.0 665.470 < 0.00000000000000022

##fractal\_dimension3 1 0.0048120 0.0064206 -5339.9 1408.911 < 0.00000000000000022

##<none>

##smoothness1 \*\*\*

##area2 \*\*\*

##concavity2 \*\*\*

##concavity3 \*\*\*

##fractal\_dimension3 \*\*\*

##---

##Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

resettest(m5, power = 2, type = "fitted", data = NONinfluenti)

##   
## RESET test  
##   
## data: m5  
## RESET = 0.089954, df1 = 1, df2 = 470, p-value 0.7644

## ETEROSCHEDASTICITÀ

Applichiamo i test di Pagan e Whyte cosi da constatare l’eteroschedasticità ,entrambi i test ci fanno accettare l’ipotesi nulla e di conseguenza di ritenere che ci sia eteroschedasticità delle varianze degli errori

# Test Pagan di eteroschedasticità  
library(lmtest)   
  
bptest(m4gam)

##   
## studentized Breusch-Pagan test  
##   
## data: m4gam  
## BP = 82.369, df = 6, p-value = 0.000000000000001157

bptest(m5)

##   
## studentized Breusch-Pagan test  
##   
## data: m5  
## BP = 7.1378, df = 5, p-value = 0.2106

# Test di WHITE   
library(car)   
ncvTest(m4gam)

## Non-constant Variance Score Test   
## Variance formula: ~ fitted.values   
## Chisquare = 133.3572, Df = 1, p = < 0.000000000000000222

ncvTest(m5)

## Non-constant Variance Score Test   
## Variance formula: ~ fitted.values   
## Chisquare = 2.334854, Df = 1, p = 0.12651

## BOOTSTRAP

L’analisi con i valori fittati è buona evidenziando che nessun valore passa per lo 0 e vedendo attraverso il plot che la distribuzione tende sempre di più alla linea obliqua

library(sandwich)  
library(psych)  
library(coefplot)

library(forestmodel)  
library("car")  
library(gvlma)  
BOOT.MOD=Boot(m5, R=1999)  
summary(BOOT.MOD, high.moments=TRUE)

##   
## Number of bootstrap replications R = 1999

## original bootBias bootSE bootMed

##(Intercept) 0.020632492 -0.00001841814 0.0006996171 0.020619016

##smoothness1 0.178772469 0.00018194464 0.0075115087 0.179013742

##area2 -0.000016373 -0.00000013565 0.0000033431 -0.000016523

##concavity2 0.122625870 0.00011356391 0.0059230285 0.122550518

##concavity3 -0.023629712 -0.00003160619 0.0008727815 -0.023661402

##fractal\_dimension3 0.334109012 0.00015240180 0.0084536635 0.333926115

## bootSkew bootKurtosis

##(Intercept) -0.1039639 0.090642

##smoothness1 -0.0196036 -0.094734

##area2 -0.0096618 0.139718

##concavity2 0.1404515 0.085519

##concavity3 -0.1052320 0.025337

##fractal\_dimension3 0.1836134 -0.023283

Confint(BOOT.MOD, level=c(.95), type="perc")

## Bootstrap percent confidence intervals

##

## Estimate 2.5 % 97.5 %

##(Intercept) 0.02063249231 0.01921422756 0.021973898091

##smoothness1 0.17877246883 0.16432074695 0.193795988453

##area2 -0.00001637331 -0.00002281004 -0.000009944816

##concavity2 0.12262586977 0.11158810635 0.134746434698

##concavity3 -0.02362971206 -0.02534795499 -0.022006913898

##fractal\_dimension3 0.33410901199 0.31869845437 0.352181767693

Confint(BOOT.MOD, level=c(.95), type="norm")

## Bootstrap normal confidence intervals

##

## Estimate 2.5 % 97.5 %

##(Intercept) 0.02063249231 0.01927968612 0.022022134779

##smoothness1 0.17877246883 0.16386823773 0.193312810651

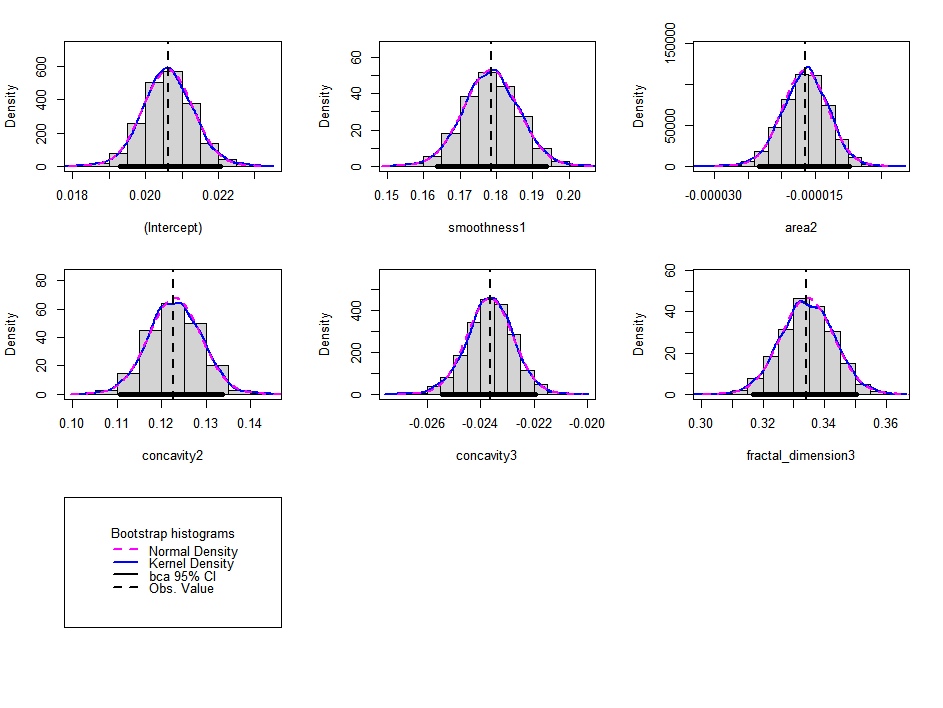
##area2 -0.00001637331 -0.00002279008 -0.000009685229

##concavity2 0.12262586977 0.11090338341 0.134121228309

##concavity3 -0.02362971206 -0.02530872610 -0.021887485653

##fractal\_dimension3 0.33410901199 0.31738773419 0.350525486205

hist(BOOT.MOD, legend="separate")



library(lmtest)  
library(sandwich)  
coeftest(m5, vcov=vcovHC(m5))

## t test of coefficients:

##

## Estimate Std. Error

##(Intercept) 0.0206324923 0.0007000594

##smoothness1 0.1787724688 0.0074100798

##area2 -0.0000163733 0.0000032834

##concavity2 0.1226258698 0.0058808387

##concavity3 -0.0236297121 0.0008727397

##fractal\_dimension3 0.3341090120 0.0084910436

## t value

##(Intercept) 29.4725

##smoothness1 24.1256

##area2 -4.9866

##concavity2 20.8518

##concavity3 -27.0753

##fractal\_dimension3 39.3484

## Pr(>|t|)

##(Intercept) < 0.00000000000000022 \*\*\*

##smoothness1 < 0.00000000000000022 \*\*\*

##area2 0.0000008652 \*\*\*

##concavity2 < 0.00000000000000022 \*\*\*

##concavity3 < 0.00000000000000022 \*\*\*

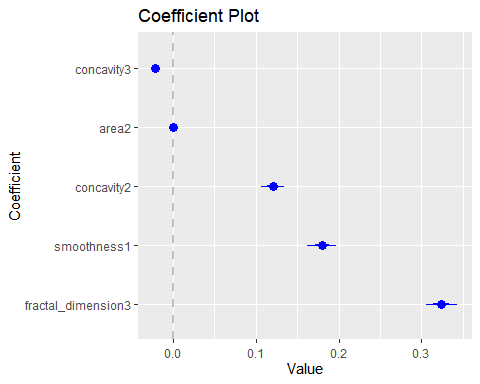
##fractal\_dimension3 < 0.00000000000000022 \*\*\*

##---

##Signif. codes:

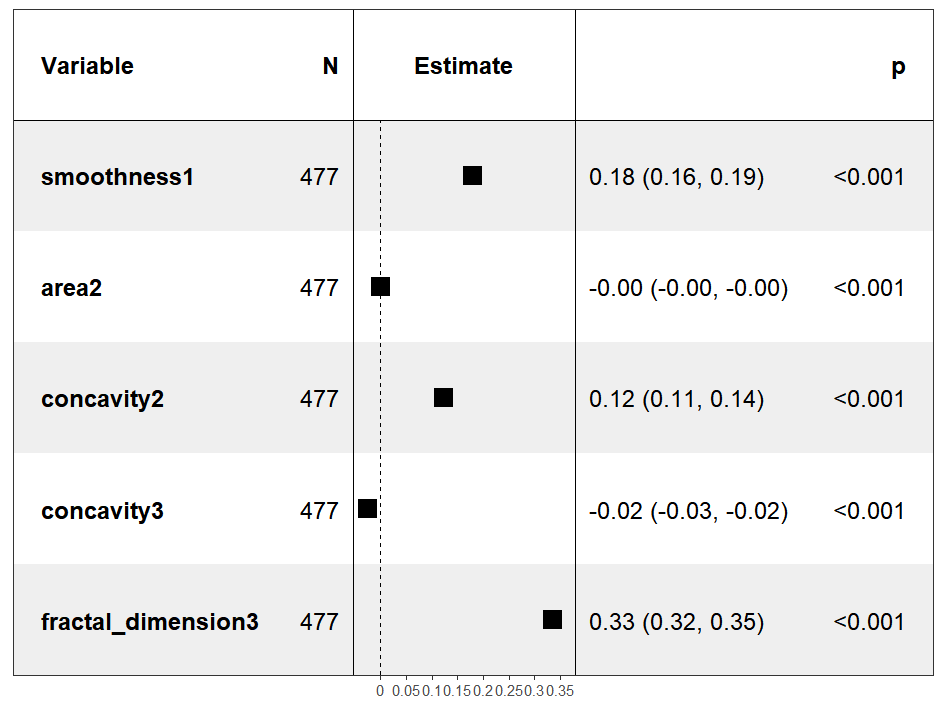
##0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

library(coefplot)  
coefplot(m5, decreasing = TRUE, sort = "magnitude",intercept=FALSE)

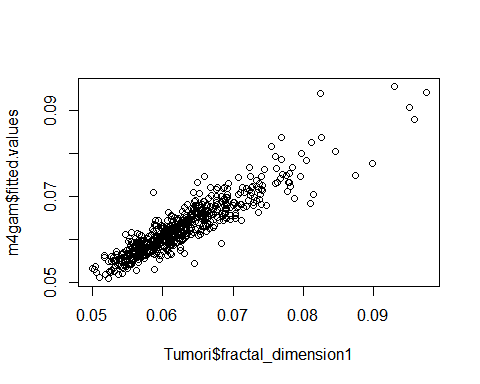


library(forestmodel)  
print(forest\_model(m5))

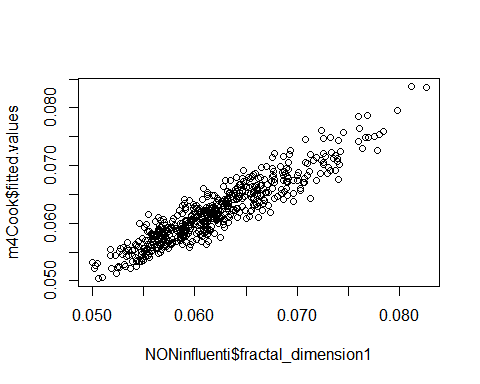
## Warning in recalculate\_width\_panels(panel\_positions, mapped\_text = mapped\_text,  
## : Unable to resize forest panel to be smaller than its heading; consider a  
## smaller text size



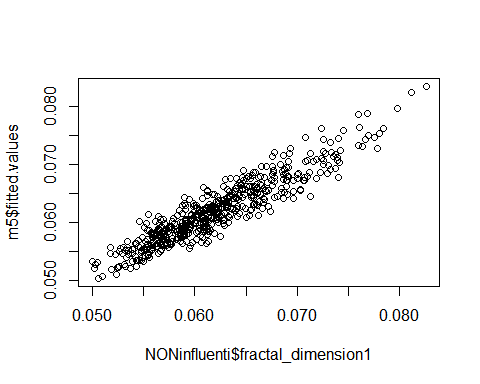
#FITTED: I VALORI SONO MIGLIORI SE SI AVVICINANO ALLA OBLIQUA  
plot(Tumori$fractal\_dimension1 , m4gam$fitted.values)



plot(NONinfluenti$fractal\_dimension1 , m4Cook$fitted.values)



plot(NONinfluenti$fractal\_dimension1 , m5$fitted.values)



summary(M3)

##Multiple R-squared: 0.8326, Adjusted R-squared: 0.8311

summary (m5)

##Multiple R-squared: 0.8892, Adjusted R-squared: 0.888

##Call:

##lm(formula = (I(fractal\_dimension1)) ~ smoothness1 + area2 +

## (area2)^2 + concavity2 + concavity3 + fractal\_dimension3,

## data = NONinfluenti)

##Residuals:

## Min 1Q Median 3Q Max

##-0.0060040 -0.0013010 -0.0001329 0.0012093 0.0049505

##Coefficients:

## Estimate Std. Error t value Pr(>|t|)

##(Intercept) 0.020632492 0.000757800 27.227 < 0.0000000000000002 \*\*\*

##smoothness1 0.178772469 0.008007590 22.325 < 0.0000000000000002 \*\*\*

##area2 -0.000016373 0.000003513 -4.661 0.00000411 \*\*\*

##concavity2 0.122625870 0.006678475 18.361 < 0.0000000000000002 \*\*\*

##concavity3 -0.023629712 0.000915997 -25.797 < 0.0000000000000002 \*\*\*

##fractal\_dimension3 0.334109012 0.008901155 37.535 < 0.0000000000000002 \*\*\*

##---

##Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

##Residual standard error: 0.001848 on 471 degrees of freedom

##Multiple R-squared: 0.8892, Adjusted R-squared: 0.888

##F-statistic: 755.8 on 5 and 471 DF, p-value: < 0.00000000000000022

## Sitografia:

Secondo alcuni studi , più le dimensioni frattali di un nucleo cellulare sono ampie, maggiore è la tendenza che quest’ultimo sia associato ad un tumore maligno.

<https://pmc.ncbi.nlm.nih.gov/articles/PMC5795727/#:~:text=Nuclear%20morphometry%20correlates%20with%20tumor,low%20and%20high%2Drisk%20groups>. (National library of Medicine)

“Nuclear morphometry correlates with tumor size, lymphnode involvement, and mitotic activity. Nuclear morphometry can be coupled with these clinicopathological features and used to prognosticate and classify breast cancer patients into low and high-risk groups.”

Dataset Utilizzato:

**Breast Cancer Wisconsin (Diagnostic)**  
dal sito (<https://archive.ics.uci.edu/ml/datasets>)

Elaborato Regressione Logistica

Marco Schiavi ; Edoardo Nada

## OBBIETTIVO DEL MODELLO:

## Il modello mira a spiegare come alcune caratteristiche del nucleo cellulare di una cellula tumorale, influenzano la sua appartenenza ad un tumore Maligno o Benigno.

## MISSING VALUE

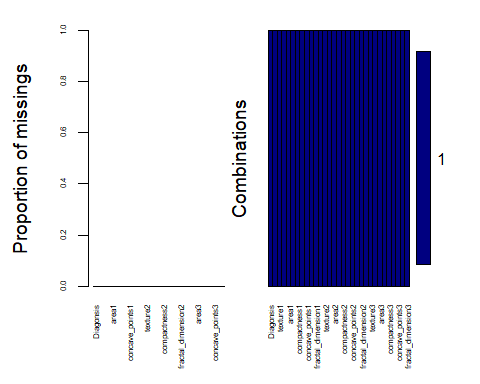
Non sono presenti valori mancanti nel dataset

library(datasets)  
library(VIM)

sapply(Tumori, function(x)(sum(is.na(x))))

## Diagonsis radius1 texture1 perimeter1   
## 0 0 0 0   
## area1 smoothness1 compactness1 concavity1   
## 0 0 0 0   
## concave\_points1 symmetry1 fractal\_dimension1 radius2   
## 0 0 0 0   
## texture2 perimeter2 area2 smoothness2   
## 0 0 0 0   
## compactness2 concavity2 concave\_points2 symmetry2   
## 0 0 0 0   
## fractal\_dimension2 radius3 texture3 perimeter3   
## 0 0 0 0   
## area3 smoothness3 compactness3 concavity3   
## 0 0 0 0   
## concave\_points3 symmetry3 fractal\_dimension3   
## 0 0 0

missingness<- aggr(Tumori, col=c('navyblue','yellow'),numbers=TRUE, sortVars=TRUE,labels=names(Tumori), cex.axis=.5,gap=2)



##   
## Variables sorted by number of missings:   
## Variable Count  
## Diagonsis 0  
## radius1 0  
## texture1 0  
## perimeter1 0  
## area1 0  
## smoothness1 0  
## compactness1 0  
## concavity1 0  
## concave\_points1 0  
## symmetry1 0  
## fractal\_dimension1 0  
## radius2 0  
## texture2 0  
## perimeter2 0  
## area2 0  
## smoothness2 0  
## compactness2 0  
## concavity2 0  
## concave\_points2 0  
## symmetry2 0  
## fractal\_dimension2 0  
## radius3 0  
## texture3 0  
## perimeter3 0  
## area3 0  
## smoothness3 0  
## compactness3 0  
## concavity3 0  
## concave\_points3 0  
## symmetry3 0  
## fractal\_dimension3 0

## CONVERSIONE VARIABILE

Converto variabile Diagonsis da stringa a factor

Tumori$Diagonsis<-factor(Tumori$Diagonsis, levels = c('M','B'), labels=c(1,0))  
  
str(Tumori)

## 'data.frame': 569 obs. of 31 variables:  
## $ Diagonsis : Factor w/ 2 levels "1","0": 1 1 1 1 1 1 1 1 1 1 ...  
## $ radius1 : num 18 20.6 19.7 11.4 20.3 ...  
## $ texture1 : num 10.4 17.8 21.2 20.4 14.3 ...  
## $ perimeter1 : num 122.8 132.9 130 77.6 135.1 ...  
## $ area1 : num 1001 1326 1203 386 1297 ...  
## $ smoothness1 : num 0.1184 0.0847 0.1096 0.1425 0.1003 ...  
## $ compactness1 : num 0.2776 0.0786 0.1599 0.2839 0.1328 ...  
## $ concavity1 : num 0.3001 0.0869 0.1974 0.2414 0.198 ...  
## $ concave\_points1 : num 0.1471 0.0702 0.1279 0.1052 0.1043 ...  
## $ symmetry1 : num 0.242 0.181 0.207 0.26 0.181 ...  
## $ fractal\_dimension1: num 0.0787 0.0567 0.06 0.0974 0.0588 ...  
## $ radius2 : num 1.095 0.543 0.746 0.496 0.757 ...  
## $ texture2 : num 0.905 0.734 0.787 1.156 0.781 ...  
## $ perimeter2 : num 8.59 3.4 4.58 3.44 5.44 ...  
## $ area2 : num 153.4 74.1 94 27.2 94.4 ...  
## $ smoothness2 : num 0.0064 0.00522 0.00615 0.00911 0.01149 ...  
## $ compactness2 : num 0.049 0.0131 0.0401 0.0746 0.0246 ...  
## $ concavity2 : num 0.0537 0.0186 0.0383 0.0566 0.0569 ...  
## $ concave\_points2 : num 0.0159 0.0134 0.0206 0.0187 0.0188 ...  
## $ symmetry2 : num 0.03 0.0139 0.0225 0.0596 0.0176 ...  
## $ fractal\_dimension2: num 0.00619 0.00353 0.00457 0.00921 0.00511 ...  
## $ radius3 : num 25.4 25 23.6 14.9 22.5 ...  
## $ texture3 : num 17.3 23.4 25.5 26.5 16.7 ...  
## $ perimeter3 : num 184.6 158.8 152.5 98.9 152.2 ...  
## $ area3 : num 2019 1956 1709 568 1575 ...  
## $ smoothness3 : num 0.162 0.124 0.144 0.21 0.137 ...  
## $ compactness3 : num 0.666 0.187 0.424 0.866 0.205 ...  
## $ concavity3 : num 0.712 0.242 0.45 0.687 0.4 ...  
## $ concave\_points3 : num 0.265 0.186 0.243 0.258 0.163 ...  
## $ symmetry3 : num 0.46 0.275 0.361 0.664 0.236 ...  
## $ fractal\_dimension3: num 0.1189 0.089 0.0876 0.173 0.0768 ...

## MULTICOLLINEARITÀ

Analizziamo la presenza di multicollinearità analizzando i plot grafici e una funzione step che lavora andando a modellare le covariate per migliorare AIC del modello, successivamente attraverso la funzione VIF siamo andando a togliere e valutare una per una le variabili fino ad arrivare al miglior modello

library(grid)  
library(colorspace)  
library(usdm)

library(psych)

library(corrgram)  
require(corrgram)  
require(car)

library(mctest)  
library(coefplot)

m1<-glm(Diagonsis~ .,data=Tumori,family = 'binomial')

## Warning: glm.fit: l'algoritmo non converge

## Warning: glm.fit: si sono verificate probabilità stimate numericamente pari a 0  
## o 1

summary(m1)

##   
## Call:  
## glm(formula = Diagonsis ~ ., family = "binomial", data = Tumori)  
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 3099896 281572 11.009 < 0.0000000000000002 \*\*\*  
## radius1 -2611364 269257 -9.698 < 0.0000000000000002 \*\*\*  
## texture1 -210634 14706 -14.323 < 0.0000000000000002 \*\*\*  
## perimeter1 -1584959 24640 -64.325 < 0.0000000000000002 \*\*\*  
## area1 139992 3907 35.828 < 0.0000000000000002 \*\*\*  
## smoothness1 164021038 8360685 19.618 < 0.0000000000000002 \*\*\*  
## compactness1 6914170 3212610 2.152 0.031382 \*   
## concavity1 -1119516 1408407 -0.795 0.426683   
## concave\_points1 18459705 5381818 3.430 0.000604 \*\*\*  
## symmetry1 -43558501 777208 -56.045 < 0.0000000000000002 \*\*\*  
## fractal\_dimension1 45543275 2168659 21.001 < 0.0000000000000002 \*\*\*  
## radius2 -35810167 1168792 -30.639 < 0.0000000000000002 \*\*\*  
## texture2 -6851608 200499 -34.173 < 0.0000000000000002 \*\*\*  
## perimeter2 -1829795 47201 -38.766 < 0.0000000000000002 \*\*\*  
## area2 687860 18351 37.483 < 0.0000000000000002 \*\*\*  
## smoothness2 -806019247 12239245 -65.855 < 0.0000000000000002 \*\*\*  
## compactness2 190764091 5732415 33.278 < 0.0000000000000002 \*\*\*  
## concavity2 -164465984 5340505 -30.796 < 0.0000000000000002 \*\*\*  
## concave\_points2 1355455618 40125123 33.781 < 0.0000000000000002 \*\*\*  
## symmetry2 -310942916 4125730 -75.367 < 0.0000000000000002 \*\*\*  
## fractal\_dimension2 -1626858724 65969621 -24.661 < 0.0000000000000002 \*\*\*  
## radius3 6595415 214299 30.777 < 0.0000000000000002 \*\*\*  
## texture3 627502 24368 25.751 < 0.0000000000000002 \*\*\*  
## perimeter3 380668 12191 31.226 < 0.0000000000000002 \*\*\*  
## area3 -96296 2741 -35.135 < 0.0000000000000002 \*\*\*  
## smoothness3 23250823 3298121 7.050 0.0000000000017928 \*\*\*  
## compactness3 -9668421 399919 -24.176 < 0.0000000000000002 \*\*\*  
## concavity3 32576904 1523449 21.384 < 0.0000000000000002 \*\*\*  
## concave\_points3 -153990201 5470934 -28.147 < 0.0000000000000002 \*\*\*  
## symmetry3 26613455 339225 78.454 < 0.0000000000000002 \*\*\*  
## fractal\_dimension3 39791264 5339652 7.452 0.0000000000000919 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 751.44 on 568 degrees of freedom  
## Residual deviance: 32006.76 on 538 degrees of freedom  
## AIC: 32069  
##   
## Number of Fisher Scoring iterations: 25

# metodo di STEP creazione modello migliore basato sull AIC  
selectedMod <- step(m1, direction="both")

m2<-glm(Diagonsis ~ radius1 + texture1 + area1 + smoothness1 + compactness1 +   
 concavity1 + concave\_points1 + symmetry1 + fractal\_dimension1 +   
 perimeter2 + area2 + smoothness2 + compactness2 + concavity2 +   
 concave\_points2 + symmetry2 + fractal\_dimension2 + radius3 +   
 texture3 + perimeter3 + area3 + concavity3 + symmetry3 +   
 fractal\_dimension3,data =Tumori,family='binomial')

## Warning: glm.fit: l'algoritmo non converge  
## Warning: glm.fit: si sono verificate probabilità stimate numericamente pari a 0  
## o 1

vif(m2)

## radius1 texture1 area1 smoothness1   
## 8611456.100 10166.407 5799699.154 1759.131   
## compactness1 concavity1 concave\_points1 symmetry1   
## 144227.263 17113.530 83514.531 2288.359   
## fractal\_dimension1 perimeter2 area2 smoothness2   
## 8116.601 78076.536 171636.174 2277.608   
## compactness2 concavity2 concave\_points2 symmetry2   
## 19895.194 216208.896 275023.729 45491.846   
## fractal\_dimension2 radius3 texture3 perimeter3   
## 43462.982 341223.458 106997.923 68525.863   
## area3 concavity3 symmetry3 fractal\_dimension3   
## 95731.768 54485.226 56932.392 47333.417

drop1(m2,test='LRT')

## Warning: glm.fit: si sono verificate probabilità stimate numericamente pari a 0  
## o 1

## Single term deletions  
##   
## Model:  
## Diagonsis ~ radius1 + texture1 + area1 + smoothness1 + compactness1 +   
## concavity1 + concave\_points1 + symmetry1 + fractal\_dimension1 +   
## perimeter2 + area2 + smoothness2 + compactness2 + concavity2 +   
## concave\_points2 + symmetry2 + fractal\_dimension2 + radius3 +   
## texture3 + perimeter3 + area3 + concavity3 + symmetry3 +   
## fractal\_dimension3  
## Df Deviance AIC LRT Pr(>Chi)   
## <none> 0.000 50.000   
## radius1 1 34.507 82.507 34.507 0.00000000424655 \*\*\*  
## texture1 1 31.204 79.204 31.203 0.00000002323591 \*\*\*  
## area1 1 32.145 80.145 32.145 0.00000001431050 \*\*\*  
## smoothness1 1 30.474 78.474 30.474 0.00000003384276 \*\*\*  
## compactness1 1 40.345 88.345 40.345 0.00000000021283 \*\*\*  
## concavity1 1 28.747 76.747 28.747 0.00000008246989 \*\*\*  
## concave\_points1 1 30.640 78.640 30.640 0.00000003105723 \*\*\*  
## symmetry1 1 29.227 77.227 29.227 0.00000006438727 \*\*\*  
## fractal\_dimension1 1 27.129 75.129 27.129 0.00000019032987 \*\*\*  
## perimeter2 1 28.714 76.714 28.714 0.00000008390982 \*\*\*  
## area2 1 32.513 80.513 32.513 0.00000001184071 \*\*\*  
## smoothness2 1 23.696 71.696 23.696 0.00000112803257 \*\*\*  
## compactness2 1 30.876 78.876 30.875 0.00000002751382 \*\*\*  
## concavity2 1 37.623 85.623 37.623 0.00000000085828 \*\*\*  
## concave\_points2 1 42.693 90.693 42.693 0.00000000006403 \*\*\*  
## symmetry2 1 29.102 77.102 29.102 0.00000006867987 \*\*\*  
## fractal\_dimension2 1 37.143 85.143 37.143 0.00000000109797 \*\*\*  
## radius3 1 28.390 76.390 28.389 0.00000009921024 \*\*\*  
## texture3 1 30.276 78.276 30.276 0.00000003747685 \*\*\*  
## perimeter3 1 25.093 73.093 25.093 0.00000054636777 \*\*\*  
## area3 1 25.061 73.061 25.060 0.00000055561544 \*\*\*  
## concavity3 1 28.973 76.973 28.973 0.00000007339871 \*\*\*  
## symmetry3 1 33.504 81.504 33.504 0.00000000711338 \*\*\*  
## fractal\_dimension3 1 32.852 80.852 32.852 0.00000000994625 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

summary(m2)

##   
## Call:  
## glm(formula = Diagonsis ~ radius1 + texture1 + area1 + smoothness1 +   
## compactness1 + concavity1 + concave\_points1 + symmetry1 +   
## fractal\_dimension1 + perimeter2 + area2 + smoothness2 + compactness2 +   
## concavity2 + concave\_points2 + symmetry2 + fractal\_dimension2 +   
## radius3 + texture3 + perimeter3 + area3 + concavity3 + symmetry3 +   
## fractal\_dimension3, family = "binomial", data = Tumori)  
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)  
## (Intercept) 5914.46 261886.32 0.023 0.982  
## radius1 6629.52 115002.01 0.058 0.954  
## texture1 -191.26 1345.37 -0.142 0.887  
## area1 -60.77 1079.21 -0.056 0.955  
## smoothness1 -39142.90 251723.69 -0.155 0.876  
## compactness1 86211.47 932557.80 0.092 0.926  
## concavity1 -28520.95 240215.54 -0.119 0.905  
## concave\_points1 -58858.72 1543927.02 -0.038 0.970  
## symmetry1 19644.82 134666.77 0.146 0.884  
## fractal\_dimension1 -162556.55 1120254.40 -0.145 0.885  
## perimeter2 1253.11 18217.26 0.069 0.945  
## area2 -156.21 2259.16 -0.069 0.945  
## smoothness2 97931.96 1472093.51 0.067 0.947  
## compactness2 -92173.51 714201.87 -0.129 0.897  
## concavity2 81310.72 1096632.56 0.074 0.941  
## concave\_points2 -439795.13 6736432.19 -0.065 0.948  
## symmetry2 103758.08 2159553.64 0.048 0.962  
## fractal\_dimension2 1092014.08 10652227.01 0.103 0.918  
## radius3 -2226.39 21338.57 -0.104 0.917  
## texture3 -72.69 3150.22 -0.023 0.982  
## perimeter3 -126.66 1355.00 -0.093 0.926  
## area3 16.26 116.50 0.140 0.889  
## concavity3 -6736.58 105090.02 -0.064 0.949  
## symmetry3 -22008.73 328258.57 -0.067 0.947  
## fractal\_dimension3 -58988.94 1031815.50 -0.057 0.954  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 751.44000538 on 568 degrees of freedom  
## Residual deviance: 0.00016713 on 544 degrees of freedom  
## AIC: 50  
##   
## Number of Fisher Scoring iterations: 25

# attraverso VIF e significatività arrivo a creare il miglior modello  
  
m5<-glm(Diagonsis~compactness1 +  
 radius2 + concavity2 +   
 texture3 + symmetry3 ,data=Tumori,family = 'binomial')  
  
drop1(m5,test='LRT')

## Single term deletions  
##   
## Model:  
## Diagonsis ~ compactness1 + radius2 + concavity2 + texture3 +   
## symmetry3  
## Df Deviance AIC LRT Pr(>Chi)   
## <none> 258.42 270.42   
## compactness1 1 310.38 320.38 51.958 0.0000000000005671 \*\*\*  
## radius2 1 394.80 404.80 136.382 < 0.00000000000000022 \*\*\*  
## concavity2 1 270.51 280.51 12.091 0.0005068 \*\*\*  
## texture3 1 328.14 338.14 69.717 < 0.00000000000000022 \*\*\*  
## symmetry3 1 284.56 294.56 26.136 0.0000003182624134 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

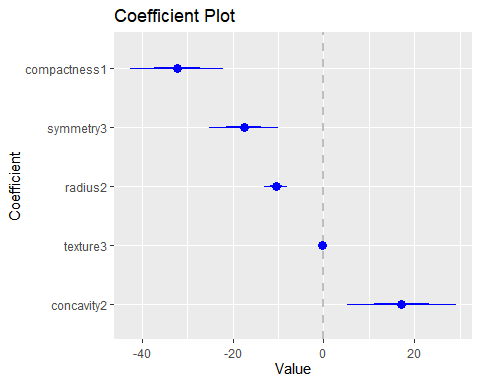
summary(m5)

##   
## Call:  
## glm(formula = Diagonsis ~ compactness1 + radius2 + concavity2 +   
## texture3 + symmetry3, family = "binomial", data = Tumori)  
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 18.97297 1.93741 9.793 < 0.0000000000000002 \*\*\*  
## compactness1 -32.34449 5.12146 -6.315 0.000000000269327528 \*\*\*  
## radius2 -10.52144 1.28357 -8.197 0.000000000000000246 \*\*\*  
## concavity2 17.14724 6.02906 2.844 0.00445 \*\*   
## texture3 -0.24012 0.03375 -7.115 0.000000000001117851 \*\*\*  
## symmetry3 -17.62667 3.82557 -4.608 0.000004073518016651 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 751.44 on 568 degrees of freedom  
## Residual deviance: 258.42 on 563 degrees of freedom  
## AIC: 270.42  
##   
## Number of Fisher Scoring iterations: 7

R=1-(258.42/751.44) #metodo calcolare omonimo R^2 nella regressione logistica  
R

## [1] 0.6561003

coefplot(m5, decreasing = TRUE, sort = "magnitude",intercept=FALSE)



## LINEARITÀ

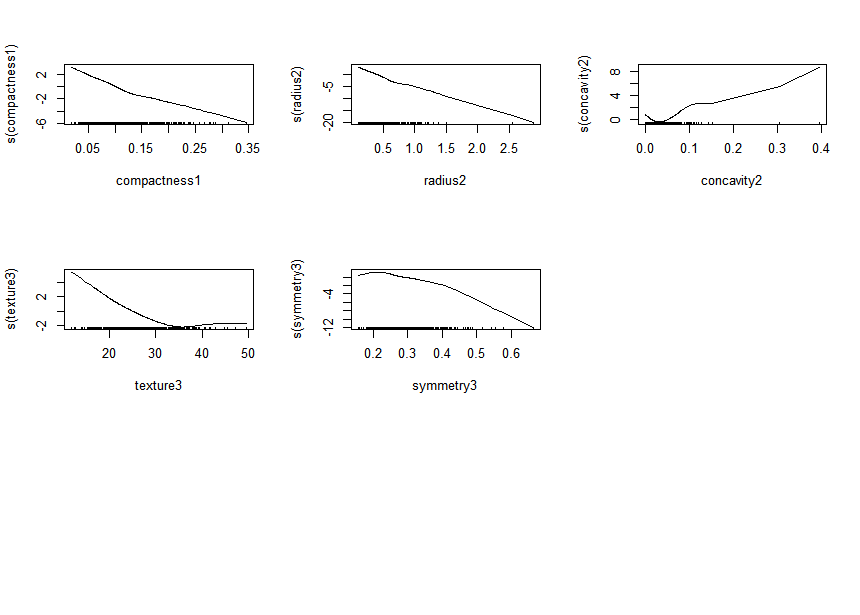
Attraverso la funzione gam e i plot decidiamo su quali variabili intervenire per cercare di linearizzarle, interveniamo solo su concavity2 essendo maggiormente significativa e avendo un chiaro plot che ci suggerisce come intervenire

library(splines)  
library(foreach)  
library(base)  
library(gam)

m5gam<- gam(Diagonsis~ s(compactness1) +  
 s(radius2) + s(concavity2) +   
 s(texture3) + s(symmetry3) ,data=Tumori,family = 'binomial')  
summary(m5gam)

##   
## Call: gam(formula = Diagonsis ~ s(compactness1) + s(radius2) + s(concavity2) +   
## s(texture3) + s(symmetry3), family = "binomial", data = Tumori)  
## Deviance Residuals:  
## Min 1Q Median 3Q Max   
## -2.67384 -0.09142 0.05899 0.30480 2.23142   
##   
## (Dispersion Parameter for binomial family taken to be 1)  
##   
## Null Deviance: 751.44 on 568 degrees of freedom  
## Residual Deviance: 223.6674 on 548.0001 degrees of freedom  
## AIC: 265.6672   
##   
## Number of Local Scoring Iterations: NA   
##   
## Anova for Parametric Effects  
## Df Sum Sq Mean Sq F value Pr(>F)   
## s(compactness1) 1 15.783 15.783 31.421 0.0000000329526805 \*\*\*  
## s(radius2) 1 28.325 28.325 56.387 0.0000000000002429 \*\*\*  
## s(concavity2) 1 5.133 5.133 10.219 0.00147 \*\*   
## s(texture3) 1 40.870 40.870 81.362 < 0.00000000000000022 \*\*\*  
## s(symmetry3) 1 13.755 13.755 27.383 0.0000002381072602 \*\*\*  
## Residuals 548 275.272 0.502   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Anova for Nonparametric Effects  
## Npar Df Npar Chisq P(Chi)   
## (Intercept)   
## s(compactness1) 3 2.6984 0.440501   
## s(radius2) 3 2.5332 0.469342   
## s(concavity2) 3 11.9662 0.007499 \*\*  
## s(texture3) 3 8.9554 0.029886 \*   
## s(symmetry3) 3 4.2467 0.236032   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

par(mfrow=c(3,3))   
plot(m5gam)  
par(mfrow=c(1,1))



m6<-glm(Diagonsis~compactness1 +  
 radius2 + concavity2 + (concavity2)^2 +  
 texture3 + symmetry3 ,data=Tumori,family = 'binomial')  
  
drop1(m6,test='LRT')

## Single term deletions  
##   
## Model:  
## Diagonsis ~ compactness1 + radius2 + concavity2 + (concavity2)^2 +   
## texture3 + symmetry3  
## Df Deviance AIC LRT Pr(>Chi)   
## <none> 258.42 270.42   
## compactness1 1 310.38 320.38 51.958 0.0000000000005671 \*\*\*  
## radius2 1 394.80 404.80 136.382 < 0.00000000000000022 \*\*\*  
## concavity2 1 270.51 280.51 12.091 0.0005068 \*\*\*  
## texture3 1 328.14 338.14 69.717 < 0.00000000000000022 \*\*\*  
## symmetry3 1 284.56 294.56 26.136 0.0000003182624134 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

summary(m6)

##   
## Call:  
## glm(formula = Diagonsis ~ compactness1 + radius2 + concavity2 +   
## (concavity2)^2 + texture3 + symmetry3, family = "binomial",   
## data = Tumori)  
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 18.97297 1.93741 9.793 < 0.0000000000000002 \*\*\*  
## compactness1 -32.34449 5.12146 -6.315 0.000000000269327528 \*\*\*  
## radius2 -10.52144 1.28357 -8.197 0.000000000000000246 \*\*\*  
## concavity2 17.14724 6.02906 2.844 0.00445 \*\*   
## texture3 -0.24012 0.03375 -7.115 0.000000000001117851 \*\*\*  
## symmetry3 -17.62667 3.82557 -4.608 0.000004073518016651 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 751.44 on 568 degrees of freedom  
## Residual deviance: 258.42 on 563 degrees of freedom  
## AIC: 270.42  
##   
## Number of Fisher Scoring iterations: 7

R=1-(258.42/751.44)  
R

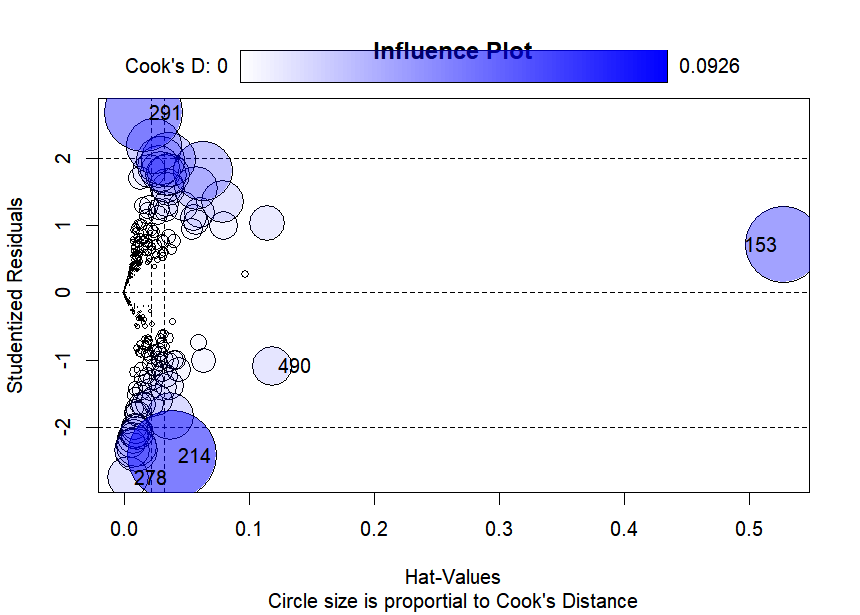
## [1] 0.6561003

## OUTLIER E LEVERAGE

Abbiamo analizzato gli Outlier e Leverage attraverso il plot e i dati ed infine abbiamo deciso di optare per una rimozione manuale di pochi valori piuttosto che utilizzare la distanza di Cook la quale andava a rimuovere valori che ritenevamo poter essere utili al modello.

library(carData)  
library(psych)  
library(dplyr)

library(car)  
influencePlot(m6, main="Influence Plot", sub="Circle size is proportial to Cook's Distance" )



## StudRes Hat CookD  
## 153 0.7142023 0.527014074 0.06767217  
## 214 -2.4221616 0.038387218 0.09261563  
## 278 -2.7531219 0.002932201 0.01999747  
## 291 2.6748435 0.015033661 0.07209543  
## 490 -1.0961033 0.118150836 0.01863036

res=data.frame(m6$residuals   
data\_used=Tumori[rownames(res),]   
cooksd <- cooks.distance(m6)  
cd=data.frame(cooksd)  
  
cutoff <- 4/(length(m6$residuals)-length(m6$coefficients)-2)  
cutoff

## [1] 0.007130125

NONinfluenti=data.frame(Tumori[cooksd < cutoff, ])

# USANDO IL LIMITE DELLA DISTANZA DI COOK TOGLIE TROPPI VALORI CHE RITENIAMO UTILI

#PREFERIAMO AGIRE MANUALMENTE E TOGLIERE SOLO 6 VALORI ANOMALI

153;209;214;291;425;505

NONinfluenti <- Tumori [-[c](http://stat.ethz.ch/R-manual/R-devel/library/base/html/c.html)(153,209,214,291,425,505),]

m7 <-glm(Diagonsis~compactness1 +  
 radius2 + concavity2 + (concavity2)^2 +  
 texture3 + symmetry3 ,data=NONinfluenti,family = 'binomial')  
  
summary(m7)

##   
## Call:  
## glm(formula = Diagonsis ~ compactness1 + radius2 + concavity2 +   
## (concavity2)^2 + texture3 + symmetry3, family = "binomial",   
## data = NONinfluenti)  
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 21.06750 2.25346 9.349 < 0.0000000000000002 \*\*\*  
## compactness1 -38.03362 6.38240 -5.959 0.000000002535710580 \*\*\*  
## radius2 -11.95644 1.47045 -8.131 0.000000000000000425 \*\*\*  
## concavity2 23.94917 10.13454 2.363 0.0181 \*   
## texture3 -0.24715 0.03603 -6.859 0.000000000006918182 \*\*\*  
## symmetry3 -21.30494 4.36685 -4.879 0.000001067367387564 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 744.79 on 562 degrees of freedom  
## Residual deviance: 232.22 on 557 degrees of freedom  
## AIC: 244.22  
##   
## Number of Fisher Scoring iterations: 7

drop1(m7,test='LRT')

## Single term deletions  
##   
## Model:  
## Diagonsis ~ compactness1 + radius2 + concavity2 + (concavity2)^2 +   
## texture3 + symmetry3  
## Df Deviance AIC LRT Pr(>Chi)   
## <none> 232.22 244.22   
## compactness1 1 284.89 294.89 52.676 0.0000000000003934383 \*\*\*  
## radius2 1 376.60 386.60 144.379 < 0.00000000000000022 \*\*\*  
## concavity2 1 241.67 251.67 9.450 0.002111 \*\*   
## texture3 1 297.47 307.47 65.252 0.0000000000000006591 \*\*\*  
## symmetry3 1 262.89 272.89 30.670 0.0000000305790438832 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

influencePlot(m7, main="Influence Plot", sub="Circle size is proportial to Cook's Distance" )

## StudRes Hat CookD  
## 113 0.9572381 0.181065488 0.02234897  
## 264 -2.4754041 0.006039606 0.01955840  
## 278 -2.8829138 0.002777913 0.02693388  
## 319 2.1324381 0.038509852 0.05123769  
## 490 -0.9049936 0.125355301 0.01249448  
## 519 2.0098114 0.043173337 0.04429687

R=1-(232.22/744.79)  
R

## [1] 0.6882074

